APPLICATION OF TIME SERIES METHODS TO RATIO ESTIMATION IN REPEATED SURVEYS

K.P. Srinath and B. Quenneville
Statistics Canada

1. INTRODUCTION

Many important surveys conducted by government and other agencies are repeated at regular intervals to provide estimates of parameters of interest and also build up a time series. Recently, there has been an emphasis on the reduction of response burden and cost in conducting these surveys. This has necessitated the investigation of the greater use of auxiliary data obtained through administrative sources to improve the efficiency of estimates based on smaller samples. One usual method of making use of auxiliary information for getting improved estimates of totals or means of characteristics of interest correlated with auxiliary variables is through ratio estimation.

In many cases, this improvement through ratio estimation is not possible right after the production of survey estimates due to a considerable time-lag in the availability of auxiliary data. In such cases, the survey estimate which is considered as preliminary is used for purposes of analysis. These preliminary estimates could undergo big revisions after the availability of auxiliary data. Therefore, there is a need to obtain improved estimates which are expected to be closer to the estimates obtained through ratio estimation later.

In this article, the problem of obtaining improved estimates of the population total Y_{T+1} for the current period using the methods of time series analysis before the availability of the auxiliary data for the current period is considered. This approach is reasonable as it is clear, that in many situations the estimate at one time period is dependent on its preceding value. It has been recognized that estimates for previous periods of time from surveys with overlapping samples do provide useful information in predicting the current population total (Smith, 1978). Several procedures for improving the simple survey estimate for the current period are considered. If the current survey estimate is denoted by Y^S_{T+1} and the estimate obtained through ratio estimation is denoted by Y^R_{T+1}, then in the first procedure described in section 2, the intention is to get the best possible predictor of Y^R_{T+1} given Y^S_{T+1}. This is done by modelling the estimates Y^R_t and the differences $Y^*_t = Y^S_t - Y^R_t$, $t = 1, 2, \ldots, T$. The models are cast in a state space form and the optimal predictor is then derived using the Kalman filter (Harvey, 1984). The method used here is similar to the one described in Rao, Srinath and Quenneville (1986).

In Section 3, sampling errors in both the estimates Y^S_{T+1} and Y^R_{T+1} are assumed. Here the intention is to get the best possible predictor of the true population total Y_{T+1}. A model on the unknown population values Y_t, $t = 1, 2, \ldots, T$ is considered. Optimal predictors of both Y_{T+1} and Y^R_{T+1} are obtained. In Section 4, it is of interest to estimate the population ratio Y_{T+1}/X_{T+1} by assuming a model on the true population ratios Y_t/X_t, $t = 1, \ldots, T$.

A procedure which involves the modelling of the auxiliary variable X_t, $t = 1, 2, \ldots, T$ is considered in Section 5. The results obtained in Section 4 are used to provide an estimate of the population total. Conditions for the optimal predictors to be better than the simple survey estimate under the sample design are derived for each case. Finally in Section 6, some concluding remarks are given.

2. PREDICTOR OF THE RATIO ESTIMATE Y^R_{T+1}

2.1 Minimum Mean Square Estimator (MMSE) of Y^R_{T+1}

We assume that the differences $Y^*_t = Y^S_t - Y^R_t$ follow a stationary AR(1) process, i.e.,

$$Y^*_t = \psi\, Y^*_{t-1} + \zeta_t, \quad t = 1, 2, \ldots, T . \tag{2.1}$$

It is also assumed that the estimate Y^R_t follow a stationary AR(1) process, i.e.,

$$Y^R_t = \phi\, Y^R_{t-1} + \varepsilon_t , \quad t = 1, \ldots, T . \tag{2.2}$$

The error vectors $(\varepsilon_t \; \zeta_t)$ are assumed to be NID $(\underset{\sim}{0}, \text{diag}\,[\sigma^2 \; \sigma_\zeta^2])$. Before the above equations are put in a state space form, the general theory of Kalman filter is briefly explained.

The state space model consists of a measurement equation

$$w_t = \underset{\sim}{z}_t' \, \underset{\sim}{\beta}_t + \xi_t, \quad t = 1, \ldots, T \tag{2.3}$$

and a transition equation

$$\underset{\sim}{\beta}_t = G_t \, \underset{\sim}{\beta}_{t-1} + \underset{\sim}{\eta}_t \, . \quad t = 1, \ldots, T \tag{2.4}$$

where $\underset{\sim}{\beta}_t$ is an mx1 state vector, $\underset{\sim}{z}_t$ is an mx1 fixed vector, G_t is a fixed mxm matrix and the errors ξ_t and $\underset{\sim}{\eta}_t$ are independent. It is further assumed that ξ_t is NID $(0, h_t)$ and $\underset{\sim}{\eta}_t$ is NID $(0, Q_t)$ where h_t is a fixed scalar and Q_t is a fixed mxm matrix.

Let $\underset{\sim}{b}_t$ be the minimum mean square estimator of $\underset{\sim}{\beta}_t$ based on all the information up to and including time t, and let P_t be the MSE matrix of $\underset{\sim}{b}_t$, i.e., the covariance matrix of $\underset{\sim}{b}_t - \underset{\sim}{\beta}_t$.

The MMSE of $\underset{\sim}{\beta}_{t+1}$ given $\underset{\sim}{b}_t$ and P_t is then given by

$$\underset{\sim}{b}_{t+1|t} = G_{t+1} \, \underset{\sim}{b}_t \tag{2.5}$$

with MSE matrix

$$P_{t+1|t} = G_{t+1} \, P_t \, G_{t+1}' + Q_{t+1} \; . \tag{2.6}$$

once w_{t+1} becomes available, this estimator of $\underset{\sim}{\beta}_{t+1}$ can be updated as follows:

$$\underset{\sim}{b}_{t+1} = b_{t+1|t} + P_{t+1|t} \, \underset{\sim}{z}_{t+1} (w_{t+1} - \underset{\sim}{z}_{t+1}' \, \underset{\sim}{b}_{t+1|t}) \, / \, f_{t+1} \tag{2.7}$$

$$P_{t+1} = P_{t+1|t} - P_{t+1|t} \, \underset{\sim}{z}_{t+1} \, \underset{\sim}{z}_{t+1}' \, P_{t+1|t} \, / \, f_{t+1} \tag{2.8}$$

$$f_{t+1} = \underset{\sim}{z}_{t+1}' \, P_{t+1|t} \, \underset{\sim}{z}_{t+1} + h_{t+1} \tag{2.9}$$

Starting values b_0 and P_0 are needed to implement the Kalman filter given by (2.5) - (2.9).

For the purpose of obtaining the optimal predictor of Y_{T+1}^R, the transition equation (2.4) can be rewritten with terms

$$\underset{\sim}{\beta}_t = \begin{vmatrix} Y_t^R \\ Y_t^* \end{vmatrix}, \quad G_t = G = \begin{vmatrix} \phi & 0 \\ 0 & \psi \end{vmatrix}, \quad \eta_t = \begin{vmatrix} \varepsilon_t \\ \xi_t \end{vmatrix}$$

and

$$Q_t = Q = \begin{vmatrix} \sigma^2 & 0 \\ 0 & \sigma_\zeta^2 \end{vmatrix} \tag{2.10}$$

as follows:

$$\begin{vmatrix} Y_t^R \\ Y_t^* \end{vmatrix} = \begin{vmatrix} \phi & 0 \\ 0 & \phi \end{vmatrix} \begin{vmatrix} Y_{t-1}^R \\ Y_{t-1}^* \end{vmatrix} + \begin{vmatrix} \varepsilon_t \\ \zeta_t \end{vmatrix} \tag{2.11}$$

The measurement equation is given by (2.3) taking $w_t = Y_t^S$, $\underset{\sim}{z}_t' = (1, \; 1)$ and $\xi_t = 0$

and is written as $\quad Y_t^S = | \; 1 \quad 1 \; | \begin{vmatrix} Y_t^R \\ Y_t^* \end{vmatrix}$

Applying the general theory of Karman filter, we let

$$\underset{\sim}{b}_t = \underset{\sim}{\beta}_t = \begin{vmatrix} Y_t^R \\ Y_t^* \end{vmatrix} \quad \text{and} \quad P_t = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} \tag{2.12}$$

It follows from (2.5) to (2.9) that the first element of MMSE of $\underset{\sim}{\beta}_{T+1}$ is given by

$$\hat{Y}_{T+1}^R = \lambda(\phi \; Y_T^R) + (1 - \lambda) (Y_{T+1}^S - \psi Y_T^*) \tag{2.13}$$

where $$\lambda = \frac{\sigma_\zeta^2}{\sigma^2 + \sigma_\zeta^2} \, .$$

The mean squared error (MSE) of $\hat{Y}_{T+1}^R$ is given by

$$\text{MSE}(\hat{Y}_{T+1}^R) = \frac{\sigma^2 \, \sigma_\zeta^2}{\sigma^2 + \sigma_\zeta^2} = \sigma^2 \, \lambda . \tag{2.14}$$

The parameters $(\phi, \; \sigma^2)$ and $(\psi, \; \sigma_\zeta^2)$ can be estimated by standard time series methods from the two series $\{Y_t^R\}$ and $\{Y_t^*\}$ $t = 1, \ldots, T$.

2.2 The bias and variance of $\hat{Y}^R_{T+1}$ under the sample design

We have $E(\hat{Y}^R_{T+1}) = E_1 E_2(\hat{Y}^R_{T+1})$ where E_2 denotes the expectation under the model and E_1 under the repeated sampling approach. It is clear that $E_2(\hat{Y}^R_{T+1}) = Y^R_{T+1}$. Therefore the bias in $\hat{Y}^R_{T+1}$ is the same as the bias in Y^R_{T+1}.

Turning to the variance of $\hat{Y}^R_{T+1}$, we have

$$V(\hat{Y}^R_{T+1}) = E_1 V_2(\hat{Y}^R_{t+1}) + V_1 E_2(\hat{Y}^R_{T+1}) .$$

Using (2.14) and under the additional assumption that the two processes Y^R_t and Y^*_t are first and second moment ergodic, we get,

$$V(\hat{Y}^{T+1}_R) = \sigma^2 \lambda + V(Y^R_{T+1}) . \qquad (2.15)$$

2.3 Comparison of the variances of $\hat{Y}^R_{T+1}$ and Y^S_{T+1} under simple random sampling

Comparing the variance of the optimal predictor $\hat{Y}^R_{T+1}$ with variance of the simple unbiased estimator Y^S_{T+1}, we have that the former is more efficient than the latter if

$$V(Y^S_{T+1}) - V(\hat{Y}^R_{T+1}) > 0 . \qquad (2.16)$$

From (2.15), we see that (2.16) implies

$$V(Y^S_{T+1}) - V(Y^R_{T+1}) > \lambda \sigma^2 \qquad (2.17)$$

Therefore (2.17) can be written as

$$\frac{V(Y^R_{T+1})}{V(Y^S_{T+1})} < \frac{1}{1+\lambda} . \qquad (2.18)$$

A condition for $\hat{Y}^R_{T+1}$ to be more efficient than Y^S_{T+1}, which is similar to the one given in Cochran (1977, p. 158) can be derived by expressing the ratio on the left hand side of (2.18) in terms of the correlation coefficient between X and Y and also the coefficients of variation of X and Y. The condition is

$$\rho_{X_T Y_T} > \frac{1}{2} \frac{C_{X_T}}{C_{Y_T}} + \left(\frac{\lambda}{1+\lambda}\right) \frac{1}{2} \frac{C_{Y_T}}{C_{X_T}} . \qquad (2.19)$$

It is seen from (2.19) that the condition for $\hat{Y}^R_{T+1}$ to be more efficient than Y^S_{T+1} is slightly stronger than the condition for the usual ratio estimator Y^R_{T+1} to be more efficient than Y^S_{T+1}. For example, if it is assumed that $C_{X_T} = C_{Y_T}$, then $\rho_{X_T Y_T}$ should be greater than .5 + (1/2) $(\lambda/1+\lambda)$. $(\lambda/1+\lambda)$ varies between 0 and 0.5. This means that $\rho_{X_T Y_T}$ should be greater than numbers between .5 and 0.75 depending on the value of λ.

3. PREDICTOR OF THE POPULATION TOTAL Y_{T+1}

3.1 Optimal predictor of Y_{T+1}

We assume that the true values Y_t, t = 1, ..., T follow a stationary AR(1) process. That is

$$Y_t = \phi Y_{t+1} + e_t , \quad t = 1, \ldots, T . \qquad (3.1)$$

As before, we also assume that the differences $Y^S_t - Y^R_t$ follow a stationary AR(1) process. That is

$$Y^*_t = \psi Y^*_{t-1} + \zeta_t , \quad t = 1, \ldots, T . \qquad (3.2)$$

e_t are assumed to be NID $(0, \sigma^2)$ and $\xi_t \sim NID(0, \sigma^2_\zeta)$ and Cov $(\zeta_t, e_t) = 0$.

The estimates Y^S_t and Y^R_t can be written as

$$Y^S_t = Y_t + u^S_t \quad \text{and} \qquad (3.3)$$

$$Y^R_t = Y_t + B(Y^R_t) + u^R_t \qquad (3.4)$$

where $B(Y^R_t)$ is the bias in the ratio estimate Y^R_t and is equal to $E(Y^R_t - Y_t)^2$, u^S_t and u^R_t are sampling errors with $E(u^S_t) = 0$, $E(u^R_t) = 0$, $V(u^R_t) = \sigma^2_R$ and $V(u^S_t) = \sigma^2_S$. Y^R_t can be expressed in terms of Y^S_t as follows.

$$Y^R_t = Y^S_t - Y^*_t = \phi Y_{t-1} - \psi Y^*_{t-1} + e_t + u^S_t - \zeta_t . \qquad (3.5)$$

In the state space form, we can write the transition equation (2.4) with terms

$$\underset{\sim}{\beta}_t = \begin{vmatrix} Y_t \\ Y_t^R \\ Y_t^* \end{vmatrix}, \quad G_t = G = \begin{vmatrix} \phi & 0 & 0 \\ \phi & 0 & -\psi \\ 0 & 0 & \psi \end{vmatrix},$$

$$\underset{\sim}{\eta}_t = \begin{vmatrix} e_t \\ e_t + u_t^S - \zeta_t \\ \zeta_t \end{vmatrix},$$

$$Q_t = Q = \begin{vmatrix} \sigma_e^2 & \sigma_e^2 & 0 \\ \sigma_e^2 & \sigma_e^2+\sigma_s^2+\sigma_\zeta^2 & -\sigma_\zeta^2 \\ 0 & -\sigma_\zeta^2 & \sigma_\zeta^2 \end{vmatrix}. \tag{3.6}$$

The measurement equation is given by (2.3) with terms

$$w_{t+1} = Y_{t+1}^S, \; z_{t+1}' = (0\ 1\ 1),$$

$$\xi_{t+1} = 0 \quad . \tag{3.7}$$

The MMSE of $\underset{\sim}{\beta}_t$, based on all the information up to and including time t, is

$$\underset{\sim}{b}_t = \begin{vmatrix} \hat{Y}_t \\ Y_t^R \\ Y_t^* \end{vmatrix} \quad \text{with MSE matrix}$$

$$P_t = \begin{vmatrix} V_2(\hat{Y}_t - Y_t) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} \tag{3.8}$$

where V_2 indicates the variance under the model.

The MMSE of $\underset{\sim}{\beta}_{t+1}$, given b_t and P_t is given by

$$b_{t+1|t} = \begin{vmatrix} \hat{Y}_{t+1|t} \\ \hat{Y}_{t+1|t}^R \\ \hat{Y}_{t+1|t}^* \end{vmatrix}$$

$$= G\underset{\sim}{b}_t = \begin{vmatrix} \phi\ \hat{Y}_t \\ \phi\ \hat{Y}_t - \psi Y^*_t \\ \psi\ Y_t^* \end{vmatrix} \tag{3.9}$$

with MSE matrix

$$P_{t+1|t} = G\ P_t\ G' + Q$$

$$= \begin{vmatrix} a_t & a_t & 0 \\ a_t & a_t+b+c & -c \\ 0 & -c & c \end{vmatrix} \tag{3.10}$$

where $a_t = \phi^2\ V_2(\hat{Y}_t - Y_t) + \sigma_e^2$,

$b = \sigma_s^2$ and $c = \sigma_\zeta^2$.

Once the survey estimate Y_{t+1}^S becomes available, the estimate $\underset{\sim}{b}_{t+1|t}$ can be updated using (2.7) and (2.8). First, from (2.9) we get $f_{t+1} = a_t + b$.

The updated estimate $\underset{\sim}{b}_{t+1}^S$ is given by

$$\hat{Y}_{t+1} = \frac{b}{a_t + b}\ \phi\ \hat{Y}_t + \frac{a_t}{a_t + b}\ Y_{t+1}^S, \tag{3.11}$$

$$\hat{Y}'^R_{t+1} = Y_{t+1}^S - \psi\ Y_t^* \tag{3.12}$$

and

$$\hat{Y}_{t+1}^* = \psi\ Y_t^* \tag{3.13}$$

with MSE matrix

$$P_{t+1}^S = \begin{vmatrix} \frac{a_t b}{a_t+b} & 0 & 0 \\ 0 & c & -c \\ 0 & -c & c \end{vmatrix}. \tag{3.14}$$

The parameters (ψ, σ_ζ^2) can be estimated by applying standard time series methods to the series $\{Y_t^*\}$, $t = 1, \ldots, T$. The parameters ϕ, σ_e^2 and σ_s^2 can be estimated using the methods outlined in Rao, Srinath and Quenneville (1986).

3.2 Comparison of the variance of the estimator $\hat{Y}'^R_{T+1}$ with that of Y_{T+1}^S

It is easy to see that the $\hat{Y}'^R_{T+1}$ has the same bias as the usual ratio estimator Y_{T+1}^R. Turning to the variance, we have

$$V(\hat{Y}'^R_{T+1}) = E_1\ V_2(\hat{Y}'^R_{T+1}) + V_1\ E_2(\hat{Y}'^R_{T+1})$$

$$= E_1\ (\sigma_\zeta^2) + V(Y_{T+1}^R)$$

$$= \sigma_\zeta^2 + V(Y_{T+1}^R) \tag{3.15}$$

We also have $\sigma_\zeta^2 = R^2\, V(X_{T+1}^S)$. Therefore

$$V(Y'^R_{T+1}) = R^2\, V(X_{T+1}^S) + V(Y_{T+1}^R) \,. \tag{3.16}$$

$\hat{Y}'^R_{T+1}$ is better than Y_{T+1}^S if

$$V(Y_{T+1}^S) - V(\hat{Y}'^R_{T+1}) > 0 \tag{3.17}$$

substituting for $V(\hat{Y}'^R_{T+1})$ in (3.17) from (3.16) and simplifying we get the condition

$$\rho_{X_{T+1}\,Y_{T+1}} > \frac{C_{X_{T+1}}}{C_{Y_{T+1}}} \tag{3.18}$$

where $C(X_{T+1})$ and $C(Y_{T+1})$ denote the coefficients of variation of X_{T+1} and Y_{T+1} respectively. This condition is much stronger than the condition for the usual ratio estimator to be better than the simple unbiased estimator. It is also to be noted that this condition for $\hat{Y}'^R_{T+1}$ is stronger than the condition for $\hat{Y}^R_{T+1}$.

Under the specific model

$$Y_t^* = \psi\, Y_{t-1}^* + \zeta_t \qquad 0 < \psi < 1 \,,$$

we have $$V(\hat{Y}'^R_{T+1}) = V(Y_{T+1}^S - \psi\, Y^*_T) \,. \tag{3.19}$$

Therefore $\hat{Y}'^R_{T+1}$ will be better than $\hat{Y}_{T+1}^S$ if

$$\rho_{Y_{T+1}^S X_T^S} > \frac{\psi}{2} R_T^S \left[\frac{V(X_T^S)}{V(Y_{T+1}^S)}\right]^{1/2} \tag{3.20}$$

where $$R_T^S = \frac{Y_T^S}{X_T^S} \,.$$

4. OPTIMAL PREDICTOR OF THE POPULATION RATIO Y_t/X_t

4.1 Estimation of the ratio

In this section though we attempt to provide an optimal predictor of the ratio Y_t/X_t, our primary interest lies in estimating the population total Y_{T+1}. We can write Y_t as $$\frac{Y_t}{X_t} \cdot X_t .$$

If R_t denotes the ratio Y_t/X_t, then we can write

$$Y_t = R_t\, X_t \qquad t = 1, \ldots, T \tag{4.1}$$

Y_t is not observed directly but only through Y_t^S. As given earlier we write

$$Y_t^S = Y_t + u_t^S \tag{4.2}$$

where u_t^S has mean zero and variance σ_S^2. Combining (4.1) and (4.2), we get

$$Y_t^S = R_t\, X_t + u_t^S \,. \tag{4.3}$$

We assume the following model for the ratio R_t:

$$R_t = R_{t-1} + e_t \,, \qquad t = 1, 2, \ldots, T \,. \tag{4.4}$$

The errors e_t are assumed to be NID $(0, \sigma_e^2)$.

In the state space form, we can write the transition equation (2.4) with terms

$$\underset{\sim}{\beta}_t = R_t \,, \quad G_t = G = 1 \,, \quad \underset{\sim}{\eta}_t = e_t \,. \tag{4.5}$$

The Kalman filter cannot be initiated at time T, since R_T is unknown. The measurement equation is given by (2.3) with terms

$$w_{t+1} = Y_{t+1}^S \,, \quad z'_{t+1} = X_{t+1} \,,$$

$$\xi_{t+1} = u_t^S \,. \tag{4.6}$$

A method of estimating R_{t+1} after the availability of Y_{t+1}^S and before the availability of X_{t+1}^S is described. The MMSE of R_t, based on all the information up to and including time t, is $\underset{\sim}{b}_t = \hat{R}_t$ with MSE $P_t = V(\hat{R}_t - R_t)$. The MMSE of R_{t+1} is then given by

$$\underset{\sim}{b}_{t+1|t} = \hat{R}_t \tag{4.7}$$

$$P_{t+1|t} = P_t + Q \tag{4.8}$$

where $Q = \sigma_e^2$.

Once Y^s_{t+1} becomes available, we can also get a survey estimate of R_{t+1} namely

$$r^s_{t+1} = \frac{Y^s_{t+1}}{X^s_{t+1}} \qquad (4.9)$$

with variance equal to $V(r^s_{t+1})$. $V(r^s_{t+1})$ is obtained in the usual manner (see Cochran, 1977). Note that the two estimates of R_{t+1}, namely, $\hat{R}_t$ and r^s_{t+1} are independent. The first one is based on all the observations up to and including time t, whereas the second one depends only on time t+1. We can express (4.7) and (4.9) as

$$\begin{vmatrix} \hat{R}_t \\ r^s_{t+1} \end{vmatrix} = \begin{vmatrix} 1 \\ 1 \end{vmatrix} R_t + \begin{vmatrix} \hat{R}_t - R_{t+1} \\ r^s_{t+1} - R_{t+1} \end{vmatrix} . \qquad (4.10)$$

The errors $\begin{vmatrix} \hat{R}_t - R_{t+1} \\ r^s_{t+1} - R_{t+1} \end{vmatrix}$

are distributed with means zero (assuming that r^s_{t+1} is approximately unbiased) and with covariance matrix

$$\begin{vmatrix} V(\hat{R}_t) & 0 \\ 0 & V(r^s_{t+1}) \end{vmatrix} . \qquad (4.11)$$

It is clear from the above set-up, that the best linear unbiased estimate of R_{t+1} is given by

$$\tilde{R}_{t+1} = \lambda \hat{R}_t + (1 - \lambda) r^s_{t+1} \qquad (4.12)$$

where $\lambda = \dfrac{V(r^s_{t+1})}{V(\hat{R}_t) + V(r^s_{t+1})}$.

Once the value of X_{T+1} becomes available, the estimate $\underset{\sim}{b}_{t+1|t}$ can be updated giving $\underset{\sim}{b}_{t+1}$ as follows

$$\underset{\sim}{b}_{t+1} = \hat{R}_{t+1} = \hat{R}_t + \frac{X_{t+1}(V(\hat{R}_t) + \sigma^2_e)}{\sigma^2_s + X^2_{t+1}(V(\hat{R}_t) + \sigma^2_e)} (Y^s_{t+1} - X_{t+1} \hat{R}_t) \qquad (4.13)$$

with MSE equal to

$$P_{t+1} = MSE(\hat{R}_{t+1}) = \frac{V(\hat{R}_t) + \sigma^2_e}{\sigma^2_s + X^2_{t+1}(V(\hat{R}_t) + \sigma^2_e)} . \qquad (4.14)$$

It is to be noted that $\hat{R}_{t+1}$ is an input into the computation of $\hat{R}_{t+2|t+1}$ and $\hat{R}_{t+2}$.

4.2 Estimation of the parameters

The procedure to obtain $\hat{R}_{t+1}$ has to be recursively applied with starting values say R_0 and P_0. The initial estimate R_0 is derived assuming that the parameter R_t is fixed for a certain period of time. This leads to the model $\beta_t = \beta_0$ and $Y^s_t = \beta_0 X_t + u^s_t$. Standard regression analysis is used to obtain an estimate b_0 of β_0 along with its estimated variance.

For estimating the variances σ^2_e and σ^2_s, we consider the maximum likelihood estimator. When the variance σ^2_e is expressed relative to σ^2_s, say $\bar{\sigma}^2_e = \sigma^2_e/\sigma^2_s$, then the log likelihood function can be written as

$$\text{Log } L(\bar{\sigma}^2_e, \sigma^2_s) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2_s - \frac{1}{2}\sum_{t=1}^{T}\log f_t - \frac{1}{2\sigma^2_s}\sum_{t=1}^{T}\frac{v^2_t}{f_t} \qquad (4.15)$$

under the normality assumption and where T is the number of observations;

$$f_{t+1} = 1 + X^2_{t+1}(V(\hat{R}_t) + \bar{\sigma}^2_e)$$

and $$v_{t+1} = Y^s_{t+1} - X_{t+1}\hat{R}_t .$$

Differentiation of (4.15) with respect to σ^2_s leads to the maximum likelihood estimate of σ^2_s:

$$\hat{\sigma}^2_s = \frac{1}{T}\sum_{t=1}^{T}\frac{v^2_t}{f_t} . \qquad (4.16)$$

It is now possible to consider the concentrated log-likelihood function

$$l_c(\bar{\sigma}^2_e) = -\frac{T}{2}\log 2\pi - \frac{T}{2} - \frac{T}{2}\log \hat{\sigma}^2 - \frac{1}{2}\sum_{t=1}^{T}\log f_t \qquad (4.17)$$

leaving out σ_s^2. The estimator of the ratio of the variance $\bar{\sigma}_e^2$ is obtained by maximizing (4.17) with numerical optimization technique. This can be done using the Fibonacci line search method assuming that $\bar{\sigma}_e^2$ lies between 0 and 1.

The latter assumption assumes that the noise in the signal R_t is less than the noise in the measurement Y_t^S.

5. A MODEL ON THE AUXILIARY VARIABLE

We assume an AR(1) model on the auxiliary variable X_t. Let the model be

$$X_t = \phi X_{t-1} + e_t , \qquad t = 1, \ldots, T \qquad (5.1)$$

We also assume a model on the differences between the survey estimate X_t^S and the true value X_t. Let this be

$$X^*_t = (X_t^S - X_t) = \psi X^*_{t-1} + \zeta_t . \qquad (5.2)$$

The error vectors $(e_t \ \zeta_t)$ are assumed to be NID $[0, \mathrm{diag}(\sigma_e^2, \sigma_\zeta^2)]$, We have,

$$X_t^S = X_t + v_t^S \qquad (5.3)$$

where v_t^S is the sampling error assumed to be independently distributed with mean zero and variance σ_v^2. We note from the previous section that

$$Y_t = R_t X_t \qquad (5.4)$$

from (4.2) we also have $Y_t^S = Y_t + u_t^S$.

In the state space form, we can write the transition equation (2.4) with terms

$$\underset{\sim}{\beta}_t = \begin{vmatrix} Y_t \\ X_t \\ X_t^* \end{vmatrix}, \quad G_t = \begin{vmatrix} 0 & R_t\phi & 0 \\ 0 & \phi & 0 \\ 0 & 0 & \psi \end{vmatrix}, \quad \eta_t = \begin{vmatrix} R_t e_t \\ e_t \\ \zeta_t \end{vmatrix},$$

$$Q_t = \begin{vmatrix} R_t^2\sigma_e^2 & R_t\sigma_e^2 & 0 \\ R_t\sigma_e^2 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_\zeta^2 \end{vmatrix} . \qquad (5.5)$$

Instead of (2.3) the measurement equation is now given by

$$\underset{\sim}{W}_t = \begin{vmatrix} Y_t^S \\ X_t^S \end{vmatrix}, \quad Z_t = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{vmatrix}, \quad \underset{\sim}{\xi}_t = \begin{vmatrix} u_t^S \\ 0 \end{vmatrix}$$

$$\text{and } H_t = \mathrm{Var}(\underset{\sim}{\xi}_t) = \begin{vmatrix} \sigma_s^2 & 0 \\ 0 & 0 \end{vmatrix} . \qquad (5.6)$$

Let $\underset{\sim}{b}_t = (\hat{Y}_t, X_t, X_t^*)$ be the MMSE of $\underset{\sim}{\beta}_t$ with MSE matrix

$$P_t = \begin{vmatrix} V(\hat{Y}_t - Y_t) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} . \qquad (5.7)$$

Using the Kalman filter with a vector of observations instead of a scalar. We can show that once the estimates Y_{t+1}^S and X_{t+1}^S become available the MMSE of Y_t is given by

$$\hat{Y}_{t+1}^{(1)} = \alpha\, Y_{t+1}^S \ \beta\, R_{t+1}\, \phi\, X_t + \gamma\, R_{t+1}\, (X_{t+1}^S - \psi\, X_t^*) \qquad (5.8)$$

where $\alpha = R_{t+1}^2\, \sigma_e^2\, \sigma_\zeta^2 / \delta$,

$\beta = \sigma_u^2\, \sigma_e^2 / \delta$,

$\gamma = \sigma_u^2\, \sigma_\zeta^2 / \delta$

and $\delta = R_{t+1}^2\, \sigma_e^2\, \sigma_\zeta^2 + \sigma_u^2\, \sigma_e^2 + \sigma_u^2\, \sigma_\zeta^2$

With MSE $\quad \mathrm{MSE}\, (\hat{Y}_{t+1}^{(1)}) = \sigma^2 \frac{\alpha}{\alpha+\beta+\gamma} \qquad (5.9)$

Once X_{t+1} become available, the MMSE of $\underset{\sim}{\beta}_{t+1}$ is given by

$$\underset{\sim}{b}_{t+1} = \begin{vmatrix} R_{t+1}\, X_{t+1} \\ X_{t+1} \\ X_{t+1}^* \end{vmatrix} \qquad (5.10)$$

with MSE matrix $P_{t+1} = 0$.

5.2 Comparison of the Variances of the Estimators $\hat{Y}^{(1)}_{t+1}$ and $\hat{Y}^{s}_{t+1}$

We have $E(\hat{Y}^{(1)}_{t+1}) = E_1 E_2(\hat{Y}^{(1)}_{t+1})$ where E_1 and E_2 denote the expectation under the design and the model respectively.

Now $E_2(\hat{Y}^{(1)}_{t+1}) = \alpha Y^{s}_{t+1} + \beta R_{t+1} X_{t+1}$

$$+ \gamma R_{t+1} (X^{s}_{t+1} - X^{*}_{t+1}). \qquad (5.11)$$

Hence $E(\hat{Y}^{(1)}_{t+1}) = Y_{t+1}$ and therefore $Y^{(1)}_{t+1}$ is unbiased for Y_{t+1}. The variance of $\hat{Y}^{(1)}_{t+1}$ si given by

$$V(\hat{Y}^{(1)}_{t+1}) = E_1 V_2 (\hat{Y}^{(1)}_{t+1}) + V_1 E_2 (\hat{Y}^{(1)}_{t+1})$$

$$= E_1 V_2(\hat{Y}^{(1)}_{t+1}) + \alpha^2 V_1 (Y^{s}_{t+1}), \qquad (5.12)$$

where

$$V_2(\hat{Y}^{(1)}_{t+1}) = \beta^2 R^2_{t+1} V_2(\phi X_t) + \gamma^2 R^2_{t+1} V(-\psi X^{*}_t)$$

$$+ 2 \beta \gamma R^2_{t+1} Cov_2 (\phi X_t, -\psi X^{*}_t) .$$

Since the processes $\{X_t\}$ and $\{X^{*}_t\}$ are independent, we have

$$V_2(\hat{Y}^{(1)}_{t+1}) = \beta^2 R^2_{t+1} \sigma^2_e + \gamma^2 R^2_{t+1} \sigma^2_\zeta . \qquad (5.13)$$

It follows from (5.12) and (5.13) that the variance of $\hat{Y}^{(1)}_{t+1}$ is given by

$$V(\hat{Y}^{(1)}_{t+1}) = \alpha^2 V_1(Y^{s}_{t+1}) + \beta^2 R^2_{t+1} \sigma^2_e$$

$$+ \gamma^2 R^2_{t+1} \sigma^2_\zeta . \qquad (5.14)$$

Now $\hat{Y}^{(1)}_{t+1}$ is better than Y^{s}_{t+1} if

$$V(Y^{s}_{t+1}) - V(\hat{Y}^{(1)}_{t+1}) > 0 .$$

That is if $(1 - \alpha^2) \sigma^2_u > \beta^2 R^2_{t+1} \sigma^2_e$

$$+ \gamma^2 R^2_{t+1} \sigma^2_\zeta \qquad (5.15)$$

substituting for α, β, and γ in (5.15) and after simplification the condition reduces to

$$((\sigma^2_u \sigma^2_e)^2 + (\sigma^2_u \sigma^2_\zeta)^2) \sigma^2_u$$

$$+ R^2_{t+1} \sigma^2_\zeta \sigma^2_e (\sigma^2_\zeta + \sigma^2_e) (\sigma^2_u)^2 > 0 . \qquad (5.16)$$

Since all the terms in (5.16) are positive, $\hat{Y}^{(1)}_{t+1}$ is always better than Y^{s}_{t+1}.

6. CONCLUDING REMARKS

If the survey estimates are available for a reasonably large number of periods previous to the current period, then it is useful to obtain "optimal" predictors of both the current population total and the ratio estimate using methods of time series analysis. The condition for the optimal predictor of the ratio estimate to be better than the simple unbiased survey estimate is slightly stronger than the condition for the usual ratio estimate to be better than the simple survey estimate. The optimal predictors of the true population total under the assumption that survey estimates are subject to sampling error are always better than the simple unbiased estimate. Though only simple models like the first order autoregressive model are considered in this paper, it should be possible to extend the results to complex time series models on all the variables including the survey errors.

REFERENCES

Cochran, W.G. (1977), Sampling Techniques, 3rd ed. New York: John Wiley & Sons.

Harvey, A.C. (1984), "A Unified View of Statistical Forecasting Procedures" Journal of Forecasting, 3, 245-275.

Rao, J.N.K., Srinath, K.P., and Quenneville, B. (1986), "Estimation of Level and Change using Current Preliminary Data", presented at the International Symposium on Panel Surveys in Washington D.C.

Smith, T.M.F. (1978), "Principles and Problems in the Analysis of Repeated Surveys", in Survey Sampling and Measurement, Ed. N.K. Namboodiri, New York: Academic Press.